

## **SYSTEMS AND METHODS FOR CONTEXTUAL MARK-UP OF FORMATTED DOCUMENTS**

The present invention claims priority to U.S. Provisional Patent Application Serial  
5 Number 60/447,379, filed February 13, 2003, the entire disclosure of which is hereby  
incorporated by reference in its entirety.

### **FIELD OF THE INVENTION**

The present invention relates to systems and methods for the contextual mark-up  
10 of formatted electronic documents. These systems and methods have many applications,  
including use for online editing and education.

### **BACKGROUND OF THE INVENTION**

The Internet has permitted users with web browsers to easily exchange  
15 information over the Internet. Educational or academic materials, for example, may be  
distributed to users for taking an educational course on-line or for facilitating grading,  
peer review, and the like. In order for the online exchange of such materials to  
successfully replace or supplement in-classroom materials, users must be provided with  
various features so that the on-line system provides certain benefits over in-classroom  
20 models. Accordingly, a need exists for various features for an on-line educational system  
and related systems.

### **DESCRIPTION OF THE DRAWINGS**

Figure 1 shows an exemplary user entry screen in some embodiments of the  
25 present invention.

Figure 2 shows an exemplary user homepage in some embodiments of the present  
invention.

Figure 3 shows an exemplary class homepage in some embodiments of the  
present invention.

30 Figure 4 shows an exemplary assignment creation page in some embodiments of  
the present invention.

Figure 5 shows an exemplary “inbox” in some embodiments of the present invention.

Figure 6 shows an exemplary markup of a formatted document in some embodiments of the present invention.

5        Figure 7 shows an exemplary marked up formatted document in some embodiments of the present invention.

Figure 8 shows an exemplary rubric library in some embodiments of the present invention.

10       Figure 9 shows an exemplary peer review markup in some embodiments of the present invention.

Figure 10 shows an exemplary peer review application in some embodiments of the present invention.

Figure 11 shows an exemplary gradebook in some embodiments of the present invention.

15       Figure 12 shows an exemplary student portfolio in some embodiments of the present invention.

Figure 13 shows exemplary information flow during the receipt of objects in some embodiments of the present invention. 1) a user submits a formatted manuscript to web servers; 2) meta-information regarding manuscript is stored into a database; 3) manuscript is saved to disk; 4 & 5) manuscript is converted into a series of images; 6) images are stored to disk; 7) user receives a confirmation that steps 1-6 were successfully completed (or that the manuscript as has been received, etc.).

20       Figure 14 shows exemplary information flow during the displaying of a document in some embodiments of the present invention. 1) reviewer requests a document to review/grade; 2) request for meta-information is sent to database; 3) meta-information regarding manuscript is returned; 4) request for document images is sent to the storage array; 5) document images are returned; 6) meta-information and document images are displayed to the user.

25       Figure 15 shows exemplary information flow during addition of marks to a document in some embodiments of the present invention. 1) user requests marked-up version of formatted document; 2) request is sent to database; 3) all mark coordinates and

any associated comments, rubrics, or editorial symbols are extracted from the database;  
4) user receives formatted document with all mark-up or review comments superimposed  
on top of the formatted document.

Figure 16 shows exemplary information flow during the displaying of marks in  
some embodiments of the present invention. 1) user requests marked-up version of  
formatted document; 2) request is sent to database; 3) all mark coordinates and any  
associated comments, rubrics, or editorial symbols are extracted from the database; 4)  
user receives formatted document with all mark-up or review comments superimposed on  
top of the formatted document.

## SUMMARY OF THE INVENTION

The present invention provides systems and methods for the contextual mark-up  
of formatted electronic documents. For example, the present invention provides systems  
and methods for adding editorial marks to fully formatted electronic documents  
exchanged over the Internet. Such systems and methods find use for on-line educational  
systems, peer review of works of authorship, and any other application where there is a  
desire to add contextual mark-ups of formatted documents. In some embodiments, the  
present invention includes systems and methods for assessing the originality of written  
materials. For example, the present invention provides systems and methods for  
detecting plagiarism of documents. The present invention further provides novel  
software and related systems to improve the use and management of formatted  
documents in an on-line environment.

For example, in some embodiments, the present invention provides a system for  
generating contextual markups (e.g., editorial comments, corrections, annotations,  
comments, etc.) of formatted documents (i.e., documents produced in a standard format,  
e.g., .pdf, .tif, etc.), comprising a processor or software configured to receive documents  
over an electronic communication network, convert unformatted documents into  
formatted documents (if necessary—e.g., if not already formatted), permit a user to add  
contextual markups to formatted documents derived from said received document to  
generate a markup document, and display said markup document on a web browser. In  
some embodiments, the formatted document is an image file (e.g., TIF, JPG, PNG, GIF,

BMP, RLE, WMF, EMF, DIB, etc.). In some embodiments, the contextual markups are added to an invisible image that maps to the image file.

In some preferred embodiments, the processor or software is further configured to assess the originality of the received document. In some embodiments, the processor or software is further configured to store information pertaining to the received document in a database (e.g., identity of submitter, grade information, editorial comments, editorial rubrics, questions, and answers).

The present invention also provides methods of using such systems. For example, the present invention provides a method for generating contextual markups of formatted documents, comprising the steps of a) providing the system above, b) inputting a document into the processor or software under conditions such that a formatted document is generated (if necessary); adding contextual markups to the formatted document to generate a markup document; and displaying the markup document.

The systems and methods of the present invention may be applied using any type of computer system, including traditional desktop computers, as well as other computing devices (e.g., calculators, phones, watches, personal digital assistants, etc.). In some embodiments, the computer system comprises computer memory or a computer memory device and a computer processor. In some embodiments, the computer memory (or computer memory device) and computer processor are part of the same computer. In other embodiments, the computer memory device or computer memory are located on one computer and the computer processor is located on a different computer. In some embodiments, the computer memory is connected to the computer processor through the Internet or World Wide Web. In some embodiments, the computer memory is on a computer readable medium (e.g., floppy disk, hard disk, compact disk, DVD, etc). In other embodiments, the computer memory (or computer memory device) and computer processor are connected via a local network or intranet.

In some embodiments, "a processor" may in fact comprise multiple processors in communication with each other for carrying out the various processing tasks required to reach the desired end result. For example, the computer of an intermediary service provider may perform some processing or information storage and the computer of a

customer linked to the intermediary service provider may perform other processing or information storage.

## **Definitions**

5           To facilitate an understanding of the present invention, a number of terms and phrases are defined below:

          The term "system" is used to refer to an on-line document management system, an example of which is described in the present specification. The term "database" is used to refer to a data structure for storing information for use by the system, and an example  
10       of such a data structure is described in the present specification.

          The term "user" refers to a person using the systems or methods of the present invention. The term "instructor" refers to a person teaching or otherwise providing content or instruction for an on-line educational system. A person may be both a user and an instructor.

15           The term "assignments" refers to work to be completed by a user for an on-line course, and examples of assignments are provided below. Assignments may include work to be completed by a user without requiring responsive information sent to an instructor, and they may also include work to be completed by a user requiring responsive information. Assignments may also include or use learning tools.

20           As used herein, the terms "processor" and "central processing unit" or "CPU" are used interchangeably and refer to a device that is able to read a program from a computer memory (*e.g.*, ROM or other computer memory) and perform a set of steps according to the program.

          As used herein, the terms "computer memory" and "computer memory device"  
25       refer to any storage media readable by a computer processor. Examples of computer memory include, but are not limited to, RAM, ROM, computer chips, digital video discs (DVD), compact discs (CDs), hard disk drives (HDD), and magnetic tape.

          As used herein, the term "computer readable medium" refers to any device or system for storing and providing information (*e.g.*, data and instructions) to a computer  
30       processor. Examples of computer readable media include, but are not limited to, DVDs, CDs, hard disk drives, magnetic tape and servers for streaming media over networks.

As used herein, the term "encode" refers to the process of converting one type of information or signal into a different type of information or signal to, for example, facilitate the transmission and/or interpretability of the information or signal. For example, image files can be converted into (*i.e.*, encoded into) electrical or digital information. Likewise, light patterns can be converted into electrical or digital information that provides an encoded video capture of the light patterns.

As used herein, the term "hyperlink" refers to a navigational link from one document to another, or from one portion (or component) of a document to another. Typically, a hyperlink is displayed as a highlighted word or phrase that can be selected by clicking on it using a mouse to jump to the associated document or documented portion.

As used herein, the term "Internet" refers to any collection of networks using standard protocols. For example, the term includes a collection of interconnected (public and/or private) networks that are linked together by a set of standard protocols (such as TCP/IP, HTTP, and FTP) to form a global, distributed network. While this term is intended to refer to what is now commonly known as the Internet, it is also intended to encompass variations that may be made in the future, including changes and additions to existing standard protocols or integration with other media (*e.g.*, television, radio, etc). The term is also intended to encompass non-public networks such as private (*e.g.*, corporate) Intranets.

As used herein, the terms "World Wide Web" or "web" refer generally to both (i) a distributed collection of interlinked, user-viewable hypertext documents (commonly referred to as Web documents or Web pages) that are accessible via the Internet, and (ii) the client and server software components which provide user access to such documents using standardized Internet protocols. Currently, the primary standard protocol for allowing applications to locate and acquire Web documents is HTTP, and the Web pages are encoded using HTML. However, the terms "Web" and "World Wide Web" are intended to encompass future markup languages and transport protocols that may be used in place of (or in addition to) HTML and HTTP.

As used herein, the term "web site" refers to a computer system that serves informational content over a network using the standard protocols of the World Wide Web. Typically, a Web site corresponds to a particular Internet domain name and

includes the content associated with a particular organization. As used herein, the term is generally intended to encompass both (i) the hardware/software server components that serve the informational content over the network, and (ii) the "back end" hardware/software components, including any non-standard or specialized components, that interact with the server components to perform services for Web site users.

As used herein, the term "HTML" refers to HyperText Markup Language that is a standard coding convention and set of codes for attaching presentation and linking attributes to informational content within documents. During a document authoring stage, the HTML codes (referred to as "tags") are embedded within the informational content of the document. When the Web document (or HTML document) is subsequently transferred from a Web server to a browser, the codes are interpreted by the browser and used to parse and display the document. Additionally, in specifying how the Web browser is to display the document, HTML tags can be used to create links to other Web documents (commonly referred to as "hyperlinks").

As used herein, the term "HTTP" refers to HyperText Transport Protocol that is the standard World Wide Web client-server protocol used for the exchange of information (such as HTML documents, and client requests for such documents) between a browser and a Web server. HTTP includes a number of different types of messages that can be sent from the client to the server to request different types of server actions. For example, a "GET" message, which has the format GET, causes the server to return the document or file located at the specified URL.

As used herein, the term "URL" refers to Uniform Resource Locator that is a unique address that fully specifies the location of a file or other resource on the Internet. The general format of a URL is protocol://machine address:port/path/filename. The port specification is optional, and if none is entered by the user, the browser defaults to the standard port for whatever service is specified as the protocol. For example, if HTTP is specified as the protocol, the browser will use the HTTP default port of 80.

As used herein, the term "in electronic communication" refers to electrical devices (e.g., computers, processors, etc.) that are configured to communicate with one another through direct or indirect signaling. For example, a conference bridge that is connected to a processor through a cable or wire, such that information can pass between the

conference bridge and the processor, are in electronic communication with one another. Likewise, a computer configured to transmit (*e.g.*, through cables, wires, infrared signals, telephone lines, etc) information to another computer or device, is in electronic communication with the other computer or device.

5           As used herein, the term "transmitting" refers to the movement of information (*e.g.*, data) from one location to another (*e.g.*, from one device to another) using any suitable means.

          As used herein, the term "XML" refers to Extensible Markup Language, an application profile that, like HTML, is based on SGML. XML differs from HTML in  
10   that: information providers can define new tag and attribute names at will; document structures can be nested to any level of complexity; any XML document can contain an optional description of its grammar for use by applications that need to perform structural validation. XML documents are made up of storage units called entities, which contain  
15   either parsed or unparsed data. Parsed data is made up of characters, some of which form character data, and some of which form markup. Markup encodes a description of the document's storage layout and logical structure. XML provides a mechanism to impose  
20   constraints on the storage layout and logical structure, to define constraints on the logical structure and to support the use of predefined storage units. A software module called an XML processor is used to read XML documents and provide access to their content and structure.

          As used herein, the term "intermediary service provider" refers to an agent providing a forum for users to interact with each other (*e.g.*, identify each other, make and receive assignments, etc). For example, an intermediary service provider may provide a forum for faculty members to create and distribute assignments to students in a  
25   class (*e.g.*, by defining the assignment and setting dates for completion), or provide a forum for students to receive and respond to assignments such as peer review assignments. The intermediary service provider also allows, for example, users to maintain a portfolio of work submitted in response to all assignments for a particular class or project and for the collection of data (such as customized questions and rubrics)  
30   which can be used to supplement knowledge base data in a library of such data. In some



embodiments, the intermediary service provider is a hosted electronic environment located on the Internet or World Wide Web.

## **DETAILED DESCRIPTION OF THE INVENTION**

5           The present invention provides system and methods for users to review, grade, comment on, and contextually mark-up a formatted manuscript, graph, image, spreadsheet, and the like, for the purposes of transmitting those reviews, grades, comments, and contextual commentaries to other individuals or groups or for archival purposes. For example, the present invention provides systems and methods for  
10   instructors who would like to grade their students' manuscripts in a digital fashion. Benefits of the systems and methods of the present invention include, but are not limited to: 1) a digital record of all manuscript comments can be maintained and tracked; 2) an author's writing progress can be tracked over time by having a computer calculate statistics regarding the rubrics or comments inserted into the manuscript; 3) temporal  
15   efficiencies are created; 4) contextual markup can supplement static commentary; and 5) reviews and grades can be normalized across multiple reviewers.

While much of the following description illustrates the invention through use of an educational environment, skilled artisans will understand that the invention is not so limited, and is generally applicable to a wide range of applications.

20           In one preferred embodiment, the systems and methods of the present invention are used by reviewers who are commenting on formatted text documents. Such uses include any that involve contextually marking of a formatted object, including, but not limited to, a grant review process, a journal review process, and an editor making changes to any type of formatted material (including images such as marketing material  
25   or music, or an accountant commenting on a spreadsheet of numbers and graphs).

A preferred embodiment of the invention is presented in the figures. For example, in one implementation, a user first enters the system using a computer in electronic communication with a processor running the system of the present invention. For example, a user authenticates themselves with a username (Fig. 1a) and password  
30   (Fig. 1b).

In the case of faculty or other user who will insert contextual markup into a formatted manuscript, the user goes to their “homepage” (Fig. 2), that lists all of the classes they are teaching (Fig. 2a). For any class, the faculty can then proceed (Fig. 2b) to the listing of all the assignments that are requested of the members (e.g., students) of that class (e.g., book reports, digital pictures, music scores, etc.) (Fig. 3).

For any particular assignment with a class (Fig. 3a), the faculty can proceed to their class “inbox” (Fig. 3b) to view all of the objects submitted to satisfy that assignment. For example, by clicking on the inbox icon (Fig. 3b), the faculty sees information regarding all of the manuscripts submitted for the assignment “Assignment 1” (Fig. 3a).

Different assignments (e.g., containers for digital object submission) can be created for a class (Fig. 3c). As part of the assignment creation, the user asks that objects be submitted beginning on a certain date (Fig. 4a) and that object submission ends on a certain date (Fig. 4b). The user also includes other pertinent information regarding object submission (Fig. 4c). Each object submitted to the system can be checked against various databases for levels of originality (Fig. 4d) and the user can select from libraries of rubrics (Fig. 4e) to use for the contextual markup of submitted works (See e.g., U.S. Pat. Publ. No. 20030164849, herein incorporated by reference in its entirety).

When the user proceeds to their “inbox” for a particular assignment (Fig. 5), they can see the various objects submitted (Fig. 5a), when they were submitted (Fig. 5b), who submitted them (Fig. 5c), an icon (e.g., color-coded to indicate originality levels) that links to the “originality report”, that documents any sections of the submitted object that were unoriginal (Fig. 5d), a link to download the object’s original file from the file system (Figs. 5e/14a), and a link to the contextual markup system (Fig. 5f).

In preferred embodiments, objects submitted to the system follow the route illustrated in Fig. 13. In embodiments, where the submitted documents are textual documents, the system converts the formatted documents into a series of images prior to contextual markup (Fig. 13, step 5). This step accounts for the lack of uniform display of formatted text (e.g., from a word processor) within a web browser. After the conversion, those images then retain all of the formatting from the word processor and can be displayed in a web browser.

In some preferred embodiments, the objects requested for contextual markup are delivered to the person who will markup those objects according to the route shown in Fig. 14.

In some embodiments, where the reviewer is marking up images and not text,  
5 there may be a bandwidth cost associated with downloading every manuscript page image every time a change is made to that page. Thus, in some embodiments, for that reason (among others—for example, see Fig. 16, section 2), the systems of the present invention allow the user to add their marks to an invisible image that maps exactly onto  
10 download each manuscript image once, because their web browser will then cache or store those images. When a mark is added to the manuscript, it appears directly on top of the manuscript image without any modifications to the manuscript image itself.

When the user selects a manuscript for contextual markup (Fig. 5f), they are sent into a contextual markup system (Fig. 6). In preferred embodiments, this system  
15 generally operates as per Fig. 15. Marks (Fig. 6e/f) are placed directly onto the formatted version of the manuscript and they are linked to commentary (Fig. 6b) and/or rubrics (Fig. 6c).

Users can then view the contextual markups for a manuscript by clicking on the markup icon from their “inbox” (Fig. 5f) and then proceeding to the commentary listing  
20 (Fig. 6n). Students or authors access contextual markup (Fig. 12b) for a submitted object (Fig. 12a) from their “portfolio” page (Fig. 12). This then takes the user to the “commentary listing” page (Fig. 7). In some preferred embodiments, users view commentary for a submitted object as outlined in Fig. 16.

Libraries of rubrics (e.g., that are used to evaluate submitted works—Figs. 4f/6c)  
25 can be managed from a “rubric library” page (Fig. 8). From this page, users can add rubrics to a particular library (Fig. 8a), delete rubrics from various libraries (Fig. 8b), move rubrics between libraries (Fig. 8c) and confirm customized symbols for rubrics (Figs. 8d/7d).

Contextual markup can also be used as an adjunct to the peer review process  
30 (Figs. 9/10). For example, when a reviewer is answering a static question (Fig. 9a), they

can refer to symbols placed in a manuscript (Figs. 9b/c) from within their answer (Fig. 9f).

When grades, scores, etc. are associated with a submitted object (Fig. 6j), in some embodiments, that information is managed and accessed via a centralized gradebook (Figs. 3d/11).

In some embodiments, the systems and methods of the present invention are provided as an application service provider (ASP) (e.g., can be accessed by users within a web-based platform via a web browser across the Internet; is bundled into a network-type appliance and run within an institution or an intranet; or is provided as a software package and used as a stand-alone system on a single computer).

Various modifications and variations of the described method and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention that are obvious to those skilled in the relevant fields are intended to be within the scope of the present invention.